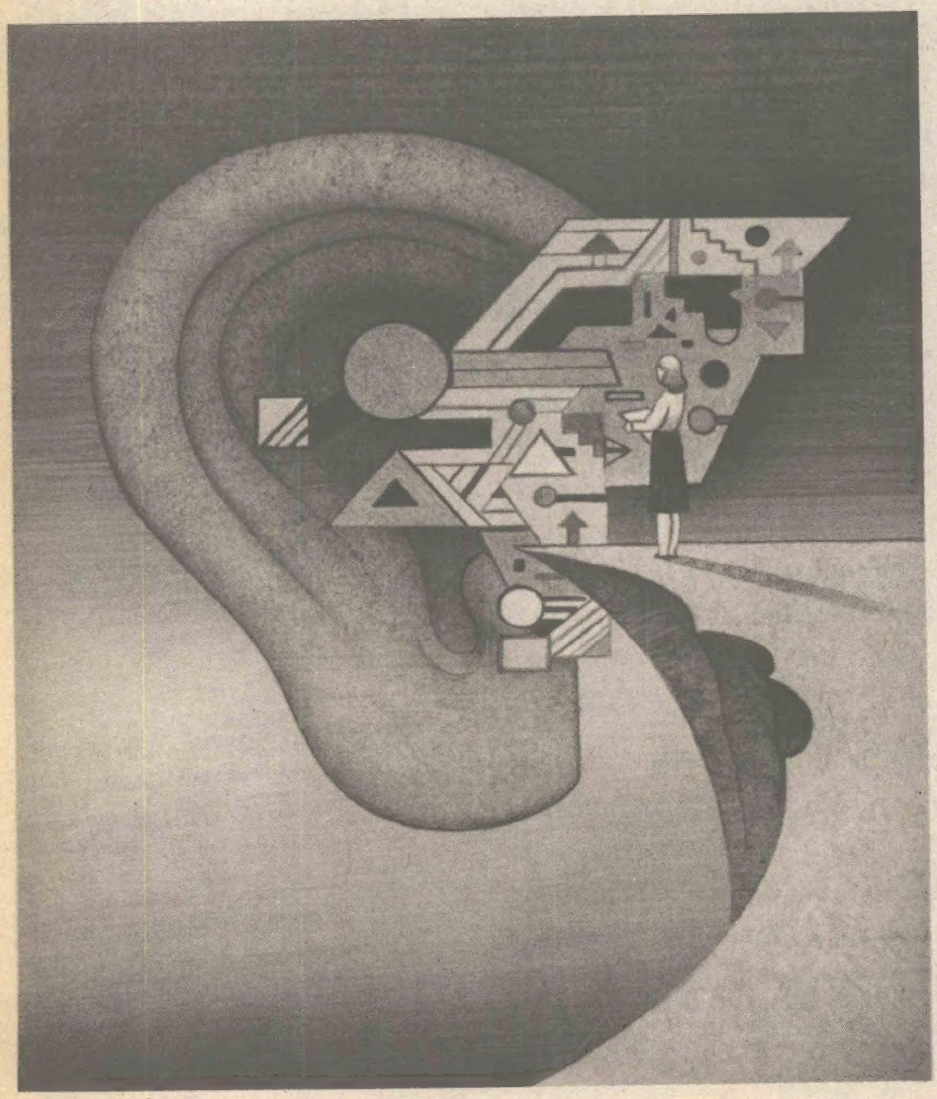
ยืน ภู่วรวรรณ

ถึงเวลาที่จะได้ยิน คอมพิวเตอร์พูด?

รายงานความเป็นไปได้ของการประยุกต์คอมพิวเตอร์ในงานโต้ตอบด้วยเสียงพูด



กว่าสิบปีมาแล้วที่คอมพิวเตอร์มีพัฒนาการเริ่มจากการเป็นหน่วยคำนวณ การประมวลผล และพัฒนาต่อมาสู่ยุคการสื่อสารข้อมูล ในยุคแรกเป็นการทำงานทางคณิตศาสตร์ คำนวณจุดของสมการที่ซับซ้อน พัฒนาการต่อมาในเรื่องการประมวลผลข้อมูล หลังจากนั้นก็มีงานประมวลผลเอกสาร ข้อความ การแสดงผล มีการแบ่งข้อมูลกันใช้ แลกเปลี่ยนข้อมูล มีการสื่อสารผ่านเครื่อง เช่น LAN และมีการส่ง E-mail ข้ามมหาสมุทร แต่จนถึงวันนี้คอมพิวเตอร์ก็ยังมีบทบาทในเรื่องการสื่อสารกับมนุษย์ได้น้อยมาก โดยเฉพาะเรื่องเสียงพูด

การจัดการเรื่องเสียงพูดเป็นเรื่องที่ต้องมีการคำนวณ ซึ่งต้องใช้ซีพียูที่เร็วมาก ในยุคแรกจึงเป็นปัญหาอยู่ แต่ปัจจุบันมีไมโครโปรเซสเซอร์พิเศษที่ใช้ในการวิเคราะห์และประมวลสัญญาณ ดังนั้นการที่จะประมวลผลเสียงพูดด้วยเวลาจริงมีแนวโน้มที่เป็นไปได้ ในที่นี้จึงได้รายงานความเป็นไปได้ของการประยุกต์คอมพิวเตอร์ในงานโต้ตอบด้วยเสียงพูด ที่จะเป็นประโยชน์ต่อผู้ใช้อย่างยิ่ง

เทคโนโลยีที่เกี่ยวกับเสียงพูด

การนำคอมพิวเตอร์มาประยุกต์ใช้ในระบบเสียงพูด พอแบ่งแยกออกเป็นสามกลุ่ม คือ การประมวลผลเนื้อหาสาระ (Content processing) การ

ควบคุมการเชื่อมต่อ (Connection Control) และการประยุกต์ในการเชื่อมต่อ (Voice interface)

การประมวลผลเนื้อหาสาระ

เป็นการนำเอาข้อมูลเสียงมาทำการประมวลผล แยกแยะหรือดำเนินการตามลักษณะการประยุกต์ใช้ การประมวลผลเนื้อหาสาระเริ่มจากการดิจิไตซ์สัญญาณเสียง เช่น ในระบบโทรศัพท์ใช้วิธีการเปลี่ยนสัญญาณเสียงให้เป็นดิจิตอลด้วย PCM (Pulse Code Modulation) เมื่อได้ PCM แล้วก็สามารถส่งต่อแบบข้อมูลดิจิตอลได้ การแปลงรหัสเป็นดิจิตอลใช้มาตรฐาน 8 บิต ดังนั้นถ้าหากต้องการส่งข้อมูลเสียงอัตราการส่ง 8 กิโลบิต ข้อมูลที่รับส่งมีความเร็ว 64 กิโลบิตต่อวินาที การใช้หลักการดิจิตอลทำให้มีความละเอียดและแม่นยำในสัญญาณเสียงมากกว่า

เมื่อได้สัญญาณเสียงในรูปตัวเลขหรือดิจิตอลแล้ว ขั้นต่อไปมีการประมวลผลทางคณิตศาสตร์ หรือที่เรียกว่า digital signal processing มีการแปลงในรูปฟังก์ชันต่าง ๆ เช่น ให้หลักการของอนุกรมฟูรีเอย์ การประมวลผลสัญญาณทำให้รู้ลักษณะหรือรูปแบบของสัญญาณเสียงพูดนั้น สามารถนำมาประยุกต์วิเคราะห์หรือสังเคราะห์สัญญาณเสียงพูดขึ้นใหม่

ควบคุมการเชื่อมต่อ

เป็นการควบคุมช่องเสียงให้ต่อกับอุปกรณ์ต่าง ๆ ตามต้องการ เช่น ในระบบชุมสายโทรศัพท์มีการควบคุมสัญญาณเสียงที่อยู่ในรูปสัญญาณดิจิตอลแล้ว มีการกำหนดสัญญาณกำหนดการเลือกช่อง เช่น DTMF (Dual Tone Multifrequency) คือสัญญาณเมื่อเรากดตัวเลขบนเครื่องโทรศัพท์ โทรศัพท์จะสร้างสัญญาณเสียงส่งตามความถี่ส่งเข้าไปในช่องสายโทรศัพท์ จากนั้นจะถอดรหัสตัวเลขเพื่อขยายการติดต่อต่อไป

การควบคุมการเชื่อมต่อมีการพัฒนาเทคนิคให้ก้าวหน้ายิ่งขึ้น จนสามารถสร้างเครือข่ายสัญญาณเสียงขนาดใหญ่ครอบคลุมได้ทั่วโลก

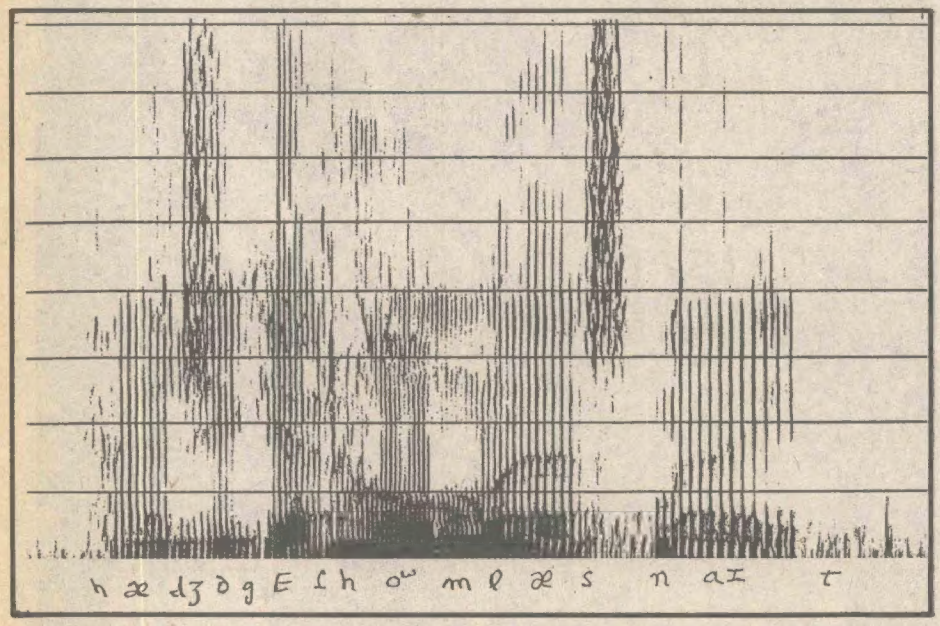
รูปที่ 1

การประยุกต์ในการเชื่อมต่อ

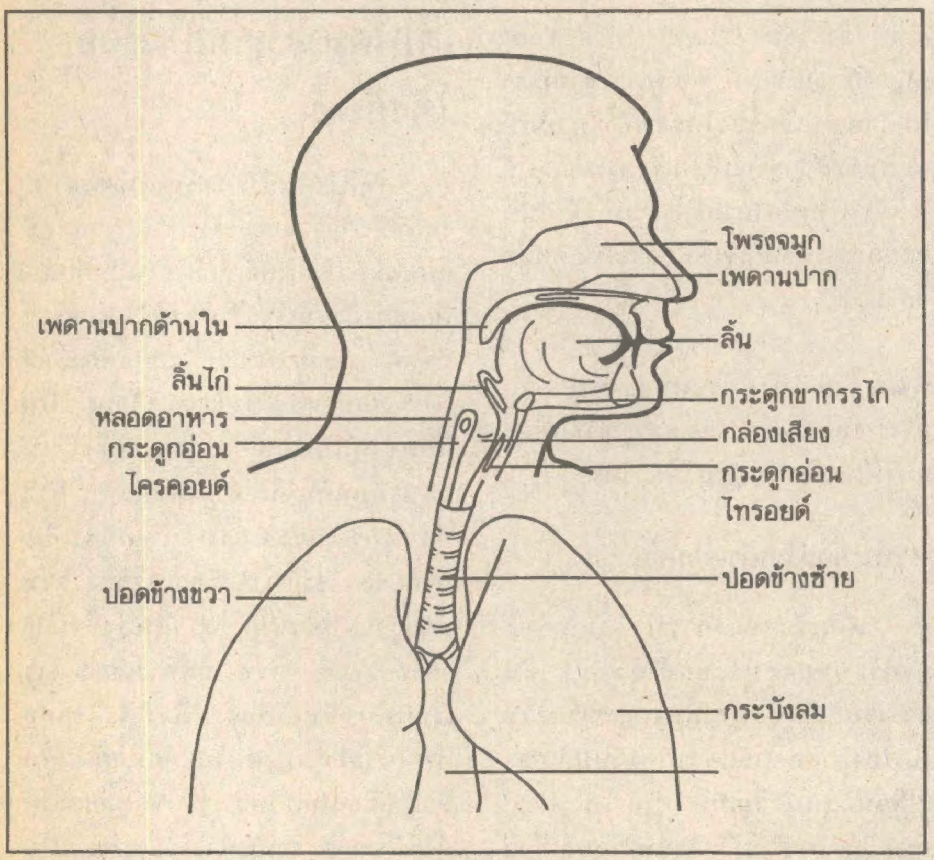
เป็นเรื่องของการนำข้อมูลเสียงมาประมวลผลประยุกต์ใช้งาน เช่น การเรียกค้นข้อมูลเสียงและนำมาใช้ในเรื่องการตอบข้อความอัตโนมัติ การเปลี่ยนตัวหนังสือเป็นคำพูด (text to speech) การให้บริการฝากข้อมูลเสียง (voice mail) การจดจำเสียงพูด ฯลฯ

เติบโตมาจากเทคโนโลยีโทรศัพท์

เทคโนโลยีโทรศัพท์เป็นเทคโนโลยีที่มีมายาวนาน เริ่มจากการสวิตช์แบบอะนาล็อก แต่เมื่อเทคโนโลยีคอมพิวเตอร์ก้าวหน้า ไมโครโปรเซสเซอร์มีขีดความสามารถประมวลผลสัญญาณได้ การแปลงสัญญาณอะนาล็อกมาเป็นดิจิตอลก็เกิดขึ้น การสวิตชิงแบบอะนาล็อกก็เปลี่ยนเป็นดิจิตอล การประยุกต์ในระบบโทรศัพท์ยังมีอีกมากมาย เช่นการประชุมหรือการรวมสัญญาณเสียงจากหลายแหล่ง (tele conference) การส่งเมล์ทางเสียง การเรียกค้นข้อมูลเสียง สิ่งที่ต้องอาศัยเทคโนโลยีขั้นสูงต่อไป ก็คือการเก็บข้อมูลเสียงในลักษณะฐานข้อมูลและให้ผู้ใช้เรียกค้น คอมพิวเตอร์จะตอบเป็นเสียงพูด การจดจำเสียงพูด



รูปที่ 2 เมื่อนำเสียงมาวิเคราะห์ทางด้านแอบความถี่และกำลังงาน จะได้สเปกตรัมที่มาใช้ประโยชน์ในการวิเคราะห์และสังเคราะห์ ในภาพเป็นการวิเคราะห์สเปกตรัมของคำพูด "How'd you get home last night"



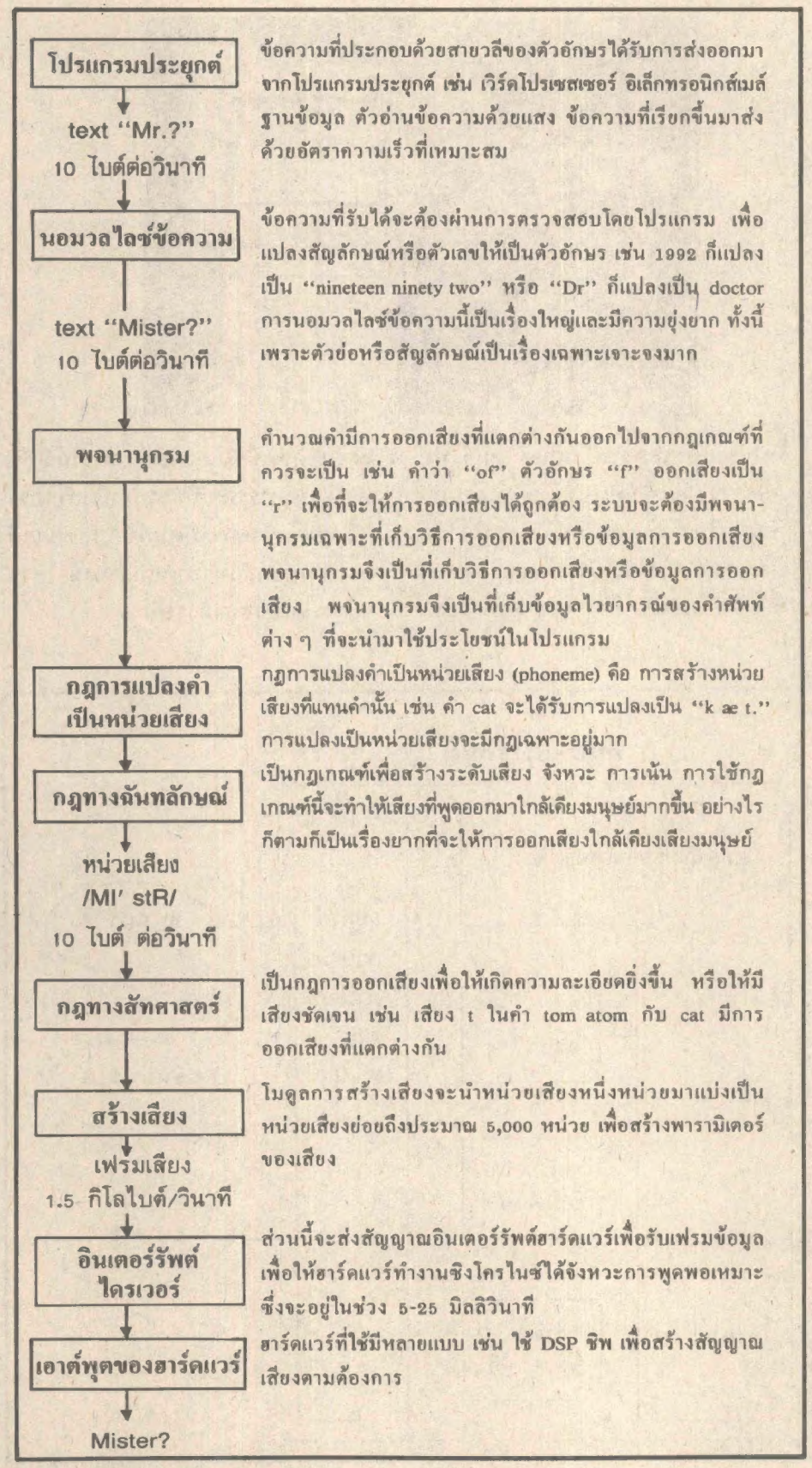
รูปที่ 3 แสดงส่วนต่าง ๆ ของอวัยวะที่มีส่วนเกี่ยวกับการสร้างเสียง

การบันทึกเสียงและการเพลย์แบค

แนวความคิดในการทำฐานข้อมูลเสียงพูดเป็นเรื่องที่พูดถึงกันมาก ทำอย่างไรจึงจะเก็บข้อมูลเสียงพูดไว้ในตัวกลาง เช่น ฮาร์ดดิสก์ ข้อมูลเสียงที่ชัดเจนและใช้งานกันได้อย่างดี มีช่วงความถี่เพียงแค่ 4 กิโลเฮิรตซ์ หากให้การสุ่มข้อมูลขนาด 8 บิต ข้อมูลเสียงที่ใช้ 1 วินาที จะมีความจุประมาณ 8 กิโลไบต์ ทั้งนี้ตามหลักการของทฤษฎีการสุ่มตัวอย่างจะต้องใช้อัตราการสุ่มไม่น้อยกว่าสองเท่าของช่วงความถี่ของสัญญาณที่ต้องการ ข้อมูลเสียงเพียง 1 วินาทีต้องใช้ความจุถึง 8 กิโลไบต์ ถ้าเป็นเช่นนี้ฐานข้อมูลเสียงจะมีมากมายมหาศาล

ด้วยหลักการของการเก็บข้อมูลในรูปตัวเลขดิจิตอลนี้เอง ผู้พัฒนาระบบจึงพยายามหาวิธีการบีบอัดข้อมูลให้เล็กลง รวมทั้งพัฒนาการของฮาร์ดแวร์เข้าช่วยในเรื่องของการสร้างสัญญาณเสียง ทำให้สัญญาณเสียงที่ใช้มีขนาดลดลงได้มาก เช่น จากหลักการของการสร้างรหัสโดยแยกเสียงออกทางด้านความถี่ที่เรียกว่า สเปกตรัม และดูในเรื่องกำลังงานของแต่ละช่วงความถี่ในขณะนั้น ทำให้สามารถลดขนาดข้อมูลเป็นรหัสลงไปได้มาก การสร้างรหัสแบบนี้จำเป็นต้องมีวงจรฮาร์ดแวร์ที่ทำงานเร็วมาก เพื่อที่จะวิเคราะห์เสียง หรือสังเคราะห์เสียงตามแถบความถี่ที่กำหนด แถบความถี่ที่แปรกับเวลานี้ที่จริงแล้วก็ใช้หลักการของอนุกรมฟูรีเอย์นั่นเอง

การสร้างโมเดลทางความถี่และกำลังงานของเสียงได้มีผู้ศึกษาและพัฒนามาใช้นานแล้ว รูปที่ 2 เป็นรูปแสดงสเปกตรัมที่ปรากฏจากการวิเคราะห์



รูปที่ 4 หลักการประมวลผลแบบ TTS

คำพูด "How'd you get home last night?" การบันทึกเสียงด้วยรหัสที่เก็บข้อมูลความถี่และกำลังงานนี้ทำให้ลดขนาดข้อมูลลงได้มาก บริษัทเท็กซัส-อินสตรูเมนต์ ผลิตชิพสำหรับการสังเคราะห์เสียงด้วยวิธีนี้ ทำให้ผู้ใช้ข้อมูลเพียงประมาณ 50 ไบต์ต่อคำพูดประมาณ 1 วินาที นับว่าวิธีการนี้จะทำให้การเก็บข้อมูลเสียงเป็นฐานข้อมูลขนาดใหญ่เป็นไปได้มาก

การวิเคราะห์และสังเคราะห์เสียงเพื่อการบันทึกและการเพลย์แบคนี้ต้องใช้โมเดลทางคณิตศาสตร์เข้าช่วย และยังต้องการขีดความสามารถเชิงคำนวณในลักษณะเฉพาะ ที่เรียกว่า การประมวลผลสัญญาณ DSP (Digital Signal Processing) การสังเคราะห์เสียงในกรณีนี้จำต้องศึกษาถึงโครงสร้างทางสรีระของมนุษย์ในการสร้างเสียงเพื่อให้ได้เสียงที่แสดงออกมาชัดเจนและเป็นธรรมชาติที่สุด โมเดลของการสร้างเสียงของมนุษย์เกี่ยวข้องกับอวัยวะส่วนต่าง ๆ อยู่มาก รูปที่ 3 เป็นรูปโมเดลส่วนที่เกี่ยวข้องกับการสร้างเสียง

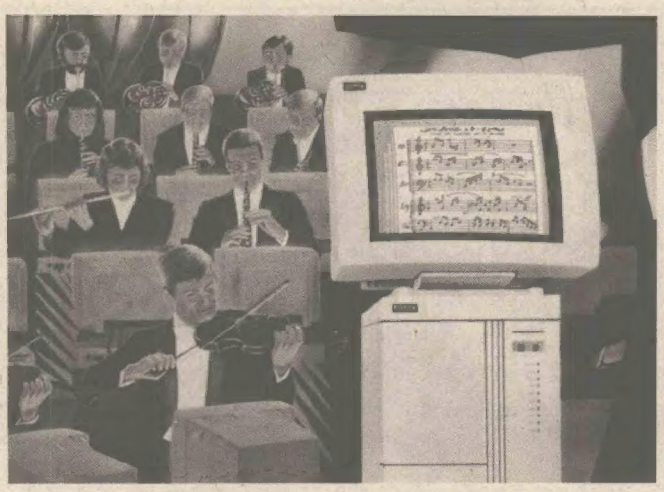
การประยุกต์ใช้ในเรื่องของการบันทึกเสียงและการเพลย์แบคมีมาก เช่น ใช้ในเครื่องตอบอัตโนมัติ ในกิจการโทรศัพท์ ใช้ในเรื่องเกม ใช้ในอุปกรณ์ไฟฟ้าที่คำพูดโต้ตอบการทำงาน นาฬิกาพูดได้ เครื่องซักผ้า กล้องถ่ายรูป เครื่องถ่ายเอกสาร เป็นต้น

การแปลงข้อความเป็นคำพูด

การแปลงข้อความเป็นคำพูดเรียกย่อ ๆ ว่า TTS (Text To Speech) การวิจัยและพัฒนาเกี่ยวกับเรื่อง TTS มีมากกว่าสามสิบปีแล้ว การศึกษาในเรื่องนี้ต้องอาศัยความรู้ทางด้านสัทศาสตร์ รูปที่ 4 แสดงหลักการใช้สำหรับการ

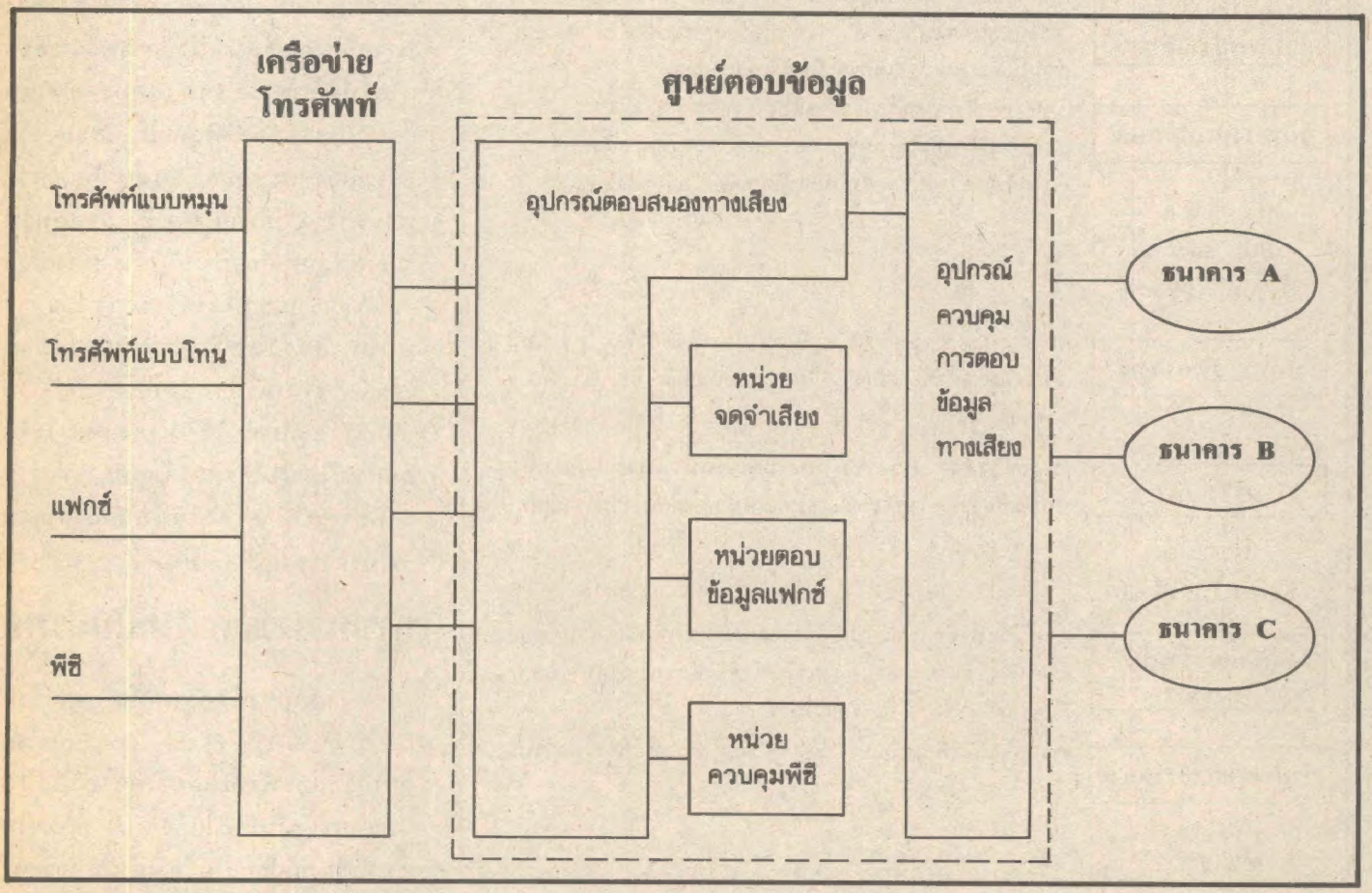
ประมวลผล TTS

ผลิตภัณฑ์การแปลงข้อความให้เป็นเสียง เป็นผลิตภัณฑ์ที่มีแนวโน้มในการนำมาประยุกต์ใช้งานได้มาก เช่น ในพจนานุกรมฉบับกระเป๋า ผู้ป่อนข้อมูลจะได้ทั้งคำอ่านและความหมาย

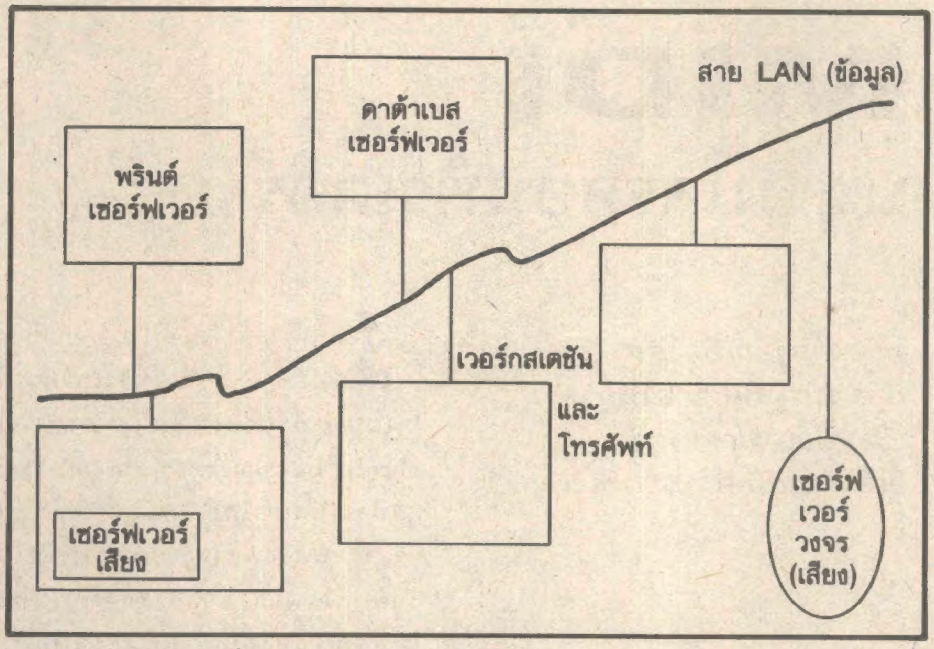


ธนาคารในประเทศญี่ปุ่นมีศูนย์ตอบข้อมูลด้วยคอมพิวเตอร์

หน่วยงานธนาคารเป็นหน่วยงานที่มีการใช้เทคโนโลยีชั้นสูง การแข่งขันการใช้เทคโนโลยีเป็นเรื่องที่ธนาคารหลายแห่งได้ดำเนินการมาแล้ว บริษัทนิปปอน เทเลกราฟ แอนด์ เทเลโฟน หรือ NTT ซึ่งเป็นบริษัทที่ดูแลระบบสื่อสารโทรคมนาคมในญี่ปุ่น จึงได้ร่วมมือกับธนาคารในญี่ปุ่นสร้างเครื่องตอบข้อมูลแบบอัตโนมัติ ศูนย์ตอบคำถามอัตโนมัติจะเชื่อมโยงกับธนาคารในญี่ปุ่นหลายธนาคาร ระบบดังกล่าวมีกลไกการสวิตช์เพื่อติดต่อกับศูนย์ข้อมูลของแต่ละธนาคารแล้วนำมาตอบคำถาม โครงสร้างระบบแสดงดังรูปที่ 5



รูปที่ 5 โครงสร้างศูนย์ตอบข้อมูลด้วยเสียงอัตโนมัติ



รูปที่ 6 การผสมเสียงเข้ากับเครือข่ายคอมพิวเตอร์

มัลติมีเดียกับเครือข่ายคอมพิวเตอร์

เมื่อเทคโนโลยีก้าวหน้ามาถึงการผสมสื่อหลายสื่อ หนทางในการผสมสื่อต่าง ๆ เข้าหากันจึงเป็นเรื่องที่มีการพัฒนาอย่างต่อเนื่อง สื่อทางเสียงเป็นสื่อที่มีการนำมาผสมโดยเริ่มจากการอัดบันทึก การเรียกค้น เนื่องจากเสียงข้อมูลรูปแบบดิจิตอลได้ จึงมีการพัฒนาให้ข้อมูลเสียงวิ่งผสมไปในเครือข่ายคอมพิวเตอร์

เครือข่ายอีเธอร์เน็ต ความเร็วของข้อมูลในการรับส่งขนาด 10 ล้านบิตต่อวินาที การนำข้อมูลเสียงใส่ลงในแพคเกจแล้วส่งต่อไปจึงเป็นเรื่องที่เป็นไปได้ ดังนั้นจึงมีผู้ออกแบบวงจรเสียงเพื่อรับเสียงส่งผ่านเข้าไปผสมร่วมกับคอมพิวเตอร์ เพื่อใช้ประยุกต์ในงานด้านต่าง ๆ โครงสร้างเครือข่ายคอมพิวเตอร์ที่ผสมระหว่างข้อมูลกับเสียง ดังรูปที่ 6

ความใฝ่ฝันอยู่ที่การให้คอมพิวเตอร์เข้าใจคำพูด

เทคโนโลยีการสร้างเสียงพูด การแปลงข้อความให้เป็นเสียงพูด เป็นเรื่องที่เป็นจริงเป็นจังแล้ว แต่การจดจำหรือเข้าใจเสียงพูดยังเป็นเรื่องของอนาคตอยู่มาก ลองจินตนาการดูว่าหากมนุษย์สามารถอ่านหนังสือให้คอมพิวเตอร์ฟัง คอมพิวเตอร์เข้าใจเสียงพูดได้ ก็สามารถแปลงเป็นตัวอักษรได้ ถ้าเป็นเช่นนี้เครื่องพิมพ์ดีดหรือแป้นพิมพ์จะหมดความหมาย การป้อนอินพุตด้วยเสียงสะดวกสบายกว่ามาก ความต้องการให้เครื่องจักรเข้าใจภาษามนุษย์ยังต้องการการค้นคว้าวิจัยอีกมาก โดยเฉพาะในเรื่องที่เกี่ยวกับความหมาย

เติมความฝัน
กับเครื่องร่อน
ที่คุณหนู
ออกแบบเอง
ด้วย...

สนุกกับ
เครื่องร่อน

สนุกสนาน เพลิดเพลินกับสิ่งประดิษฐ์ที่ง่ายแสนง่ายมีคำอธิบายพร้อมภาพประกอบ เพียงหาอุปกรณ์ใกล้ตัวก็สนุกได้หลายรูปแบบของเครื่องร่อน ร่มชูชีพ และว่าว พร้อมบทความเกี่ยวกับวิวัฒนาการการบิน การกำเนิดร่มชูชีพ ประวัติความเป็นมาของว่าว และเสริมความรู้เกี่ยวกับวิทยาศาสตร์ทางการบิน

หรือสั่งซื้อโดยตรงได้ที่
บริษัท ซีเอ็ดยูเคชั่น จำกัด
800/43-45 ซอยตระกูลสุข ถนนอโศก-ดินแดง ห้วยขวาง กรุงเทพฯ 10400 โทร. 248-6280-9 (เวลาทำการ จันทร์-ศุกร์ 8.10-17.30 น.) โดยส่งธนาณัติสั่งจ่าย ปท.ดินแดง ในนามบริษัทฯ หรือเช็คส่วนตัวของธนาคารในเขตกรุงเทพฯ พร้อมค่าจัดส่ง

ค่าจัดส่ง สั่งซื้อไม่เกิน 100 บาท ค่าจัดส่ง 8 บาท สั่งซื้อเกิน 100 บาท ค่าจัดส่ง 8%

พิเศษ สำหรับสมาชิกวารสารของซีเอ็ด สั่งซื้อโดยตรงที่บริษัทฯ ได้ในราคาลด 15% เพิ่มค่าลงทะเบียน 4 บาท ค่าจัดส่งฟรี โดยระบุหมายเลขสมาชิกด้วย